数理解析研究所講究録
167巻 1972年 103-131

A Geometry of Parameter Space and it's Statistical Interpretation

Tadashi Yoshizawa
Yamanashi University

0. Abstract

The space which is composed of the parameters of a distribution, the parameter space, may be considered as a Riemannian space by introducing an appropriate metric under some conditions. Invariant quantities in a geometry have very important meanings in its application to various fields of science. It is shown that a necessary and sufficient condition for existence of covariance stabilizing transformation is that the Riemann-Christoffel curvature tensor calculated from the metric is zero. Some population spaces with constant Gaussian curvatures which are immersed in higher dimensional Euclidean spaces and the concept of distance in population spaces are discussed with examples. Finally the relations between this geometry and Fisher's information matrix or the other definitions of distance or divergence between two distributions are mentioned.

1. Introduction

In this paper we consider a parameter space as a Riemannian space by introducing a fundamental tensor of the metric and discuss the statistical meanings of various invariant quantities in the Riemannian space. The inverse of a covariance matrix of a real-valued random vector whose asymptotic distribution is a Normal distribution is used as a fundamental tensor of the metric in section 2.

In section 3, the Riemann-Christoffel curvature tensor which is a typical invariant quantity in a Riemannian space is interpreted. It is shown that the condition for existence of a covariance stabilizing transformation is that all components of the Riemann-Christoffel curvature tensor are zeros, and that a condition given by Holland (1971) in two dimensional case is equivalent to our condition.

In section 4, by calculating Gaussian curvatures of some parameter spaces, it is shown that parameter spaces of a multinomial distribution and one dimensional normal distribution have positive and negative constant Gaussian curvatures respectively. Since a Riemannian space with a constant curvature may be interpreted as a fundamental hyperquadric of a higher (by one) dimensional Euclidean space, the coordinates of the Euclidean space in which the Riemannian space is immersed are given.

In section 5, it is shown that a statistic with asymptotically constant variance is formed by a transformation associated with geodesics in the parameter space. Such transformations are given with respect to the same examples as in section 4. Finally in section 6, the relation between this geometry and Fisher's information matrix is discussed.

Details of tiresome calculations will be given in the appendices.

The idea that a parameter space may be regarded as a Riemannian space by introducing an appropriate metric has been proposed by Rao (1945) and Yoshizawa (1962) using Fisher's information matrix. Rao gave a solution for geodesics in the case of a parameter space composed of the parent mean and standard deviation of a normal distribution in more complicated form than ours and tried to use it for testing in large samples without noticing its asymptotically constant variance. Yoshizawa, at the suggestion of Professor Moriguti, gave some examples of spaces with constant Gaussian curvatures and discussed their statistical meanings.

Recently Holland (1971) considered an asymptotic concept of a covariance stabilizing transformation and gave a necessary and sufficient condition for its existence. The author studied this problem again receiving impetus from Holland's paper.

2. Introduction of a fundamental tensor of the metric

Let x_n and θ be a real valued p-dimensional random vector with coordinates x_n^i and a p-dimensional parameter with coordinates θ^i, respectively. This vector parameter, θ, varies over D, an open, simply connected region of R^p. Finally, assume that $\sqrt{n}(x_n-\theta)$ has an asymptotic multivariate Normal distribution with zero mean vector and a non-singular covariance matrix, i.e.

(1) $$\pounds[\sqrt{n}(x_n-\theta)] \to N(0,\Sigma(\theta))$$

where $\Sigma(\theta)$ is positive definite for all $\theta \in D$.

Let θ' be a new p-dimensional coordinate system transformed one to one by

(2) $$\theta' = f(\theta).$$

It is easily seen that $\sqrt{n}(f(x_n)-f(\theta))$ has an asymptotic multivariate Normal distribution with zero mean vector and a non-singular covariance matrix, i.e.

(3) $$\pounds\sqrt{n}[f(x_n)-f(\theta)] \to N(0,\Sigma'(\theta))$$

where

(4) $$\Sigma'(\theta) = (\frac{\partial\theta'}{\partial\theta})\Sigma(\theta)(\frac{\partial\theta'}{\partial\theta})$$

and $(\frac{\partial\theta'}{\partial\theta})$ means the Jacobian matrix of the transformation (2) under conditions that all partial derivatives $\partial\theta'^i/\partial\theta^k$ exist and that the Jacobian is not zero (see Holland [1971]).

Let g^{ij} and g'^{ij} denote the elements of $\Sigma(\theta)$ and $\Sigma'(\theta)$ respectively. We may rewrite (4) as

$$g'^{ij} = g^{k\ell} \frac{\partial\theta'^i}{\partial\theta^k} \frac{\partial\theta'^j}{\partial\theta^\ell} \tag{5}$$

where Einstein summation convention* is used. We shall use this convention hereafter. From (5) it is seen that g^{ij} is a contravariant tensor of the second order.

Let g_{ij} be components of the inverse matrix $\Sigma(\theta)$, i.e.

$$g^{ij} g_{kj} = \delta^i_k \tag{6}$$

where δ^i_k's are Kronecker deltas. Since it is seen from (5) and (6) that the law of transformation of g_{ij} to g'_{ij} is

$$g'_{ij} = g_{k\ell} \frac{\partial\theta^k}{\partial\theta'^i} \frac{\partial\theta^\ell}{\partial\theta'^j} \tag{7}$$

g_{ij} is a covariant tensor. It is also positive definite.

Therefore we may take formally as the basis of the metric of a parameter space, a space of parameter θ, a real fundamental quadratic form

$$\phi = g_{ij} d\theta^i d\theta^j \ . \tag{8}$$

The tensor g_{ij} is called the fundamental tensor of the metric. If element of length ds is defined by

$$ds^2 = g_{ij} d\theta^i d\theta^j$$

noticing that g_{ij} is positive definite, from (7) it is seen that ds^2 is invariant under the transformation (2). This definition of ds may be acceptable as an extension of the concept of concentration matrix by Dempster [1969]. When θ

---

* When the same letter appears in any terms as a subscript and superscript, it is understood that this letter is summed up for all the values, say p, which this letter takes. k and ℓ in (5) are the examples of such letters, called dummy indecies.

is a mean vector of a multivariate Normal distribution with constant Σ this distance is no more than Mahalanobis' generalized distance (Mahalanobis [1936]).

Invariant quantities in a geometry have very important meanings in its application. In the following chapters we will discuss statistical meanings of several geometrical concepts in parameter space as a Riemannian space.

3. Riemann-Christoffel curvature tensor and covariance stabilizing transformation

Holland [1971] defined a covariance stabilizing transformation f as a set of functions (2) which satisfies

$$g^{ij} \frac{\partial \theta'^{k}}{\partial \theta^{i}} \frac{\partial \theta'^{\ell}}{\partial \theta^{j}} = \delta^{k}_{\ell} \; . \tag{10}$$

We may rewrite this condition as

$$g_{ij} \frac{\partial \theta^{i}}{\partial \theta'^{k}} \frac{\partial \theta^{j}}{\partial \theta'^{\ell}} = \delta^{k}_{\ell} \tag{11}$$

from (6).

A space which has δ^{k}_{ℓ} as the fundamental tensor of the metric is called a Euclidean space. Therefore the condition (10) or (11) for existence of a covariance stabilizing transformation may be replaced by the condition that a Riemannian space be Euclidean. Since it is well known that the latter condition is that the Riemann-Christoffel curvature tensor vanishes(e.g. see Veblen [1933], pp 69-71), we obtain the following theorem.

Theorem 1: The necessary and sufficient condition that a covariance stabilizing transformation exists is that all components of the Riemann-Christoffel curvature tensor are equal to zero, i.e.

$$R_{hijk} = 0, \tag{12}$$

where

$$R_{hijk} = \frac{\partial}{\partial\theta^j}[ik,h]-\frac{\partial}{\partial\theta^k}[ij,h]+\{{}^{\ell}_{ij}\}[hk,\ell]-\{{}^{\ell}_{ik}\}[hj,\ell], \tag{13}$$

$$[ij,k] = \frac{1}{2}\left(\frac{\partial g_{ik}}{\partial\theta^j}+\frac{\partial g_{jk}}{\partial\theta^i}-\frac{\partial g_{ij}}{\partial\theta^k}\right), \tag{14}$$

$$\{{}^{\ell}_{ij}\} = g^{\ell k}[ij,k]. \tag{15}$$

$[ij,k]$ and $\{{}^{\ell}_{ij}\}$ are called Christoffel 3-index symbols of the first and second kinds, respectively.

From the definition (6) we find that the components of Riemann-Christoffel curvature tensor satisfy the following identities:

$$\begin{aligned} R_{hijk} &= -R_{ihjk}, \\ R_{hijk} &= -R_{hikj}, \\ R_{hijk} &= R_{jkhi}, \end{aligned} \tag{16}$$

and

$$R_{hijk} + R_{hjki} + R_{hkij} = 0. \tag{17}$$

The number of independent components of Riemann-Christoffel curvature tensor is at most $p^2(p^2-1)/12$ due to the identities (16) and (17). In the two dimensional case

$$R_{1212} = R_{2121} = -R_{1221} = -R_{2112}$$

and the other components are zero. Therefore we may write the condition (12) in two dimensional case as

$$R_{1212} = 0. \tag{18}$$

The equivalence between (18) and the condition obtained by Holland [1971] will be shown in Appendix A.

4. Gaussian curvature and space with a constant curvature

Gaussian curvature defined as

$$K = \frac{R_{hijk}}{g_{hj}g_{ik}-g_{hk}g_{ij}} \tag{19}$$

has an important role in Riemannian geometry. In particular a p-dimensional Riemannian space with constant Gaussian curvature can be immersed in p+1 dimensional Euclidean space and can be interpreted as a fundamental hyperquadric of the Euclidean space. The hyperquadric is defined by

$$\sum_{\alpha=1}^{p+1} c_\alpha (z^\alpha)^2 = \frac{1}{K}, \tag{20}$$

where c_α's are plus or minus one according to the character of the fundamental form and z^α's are a set of solutions of the equations

$$\frac{\partial^2 z^\alpha}{\partial\theta^i \partial\theta^j} - \frac{\partial z^\alpha}{\partial\theta^h}\{^{\,h}_{ij}\} = -Kg_{ij}z^\alpha, \quad (\alpha=1,\ldots,p+1) \tag{21}$$

(See Eisenhart [1926]). The left-hand of the above equation (21) is called covariant differentiation with respect to a tensor g_{ij} and denoted by $z^\alpha{}_{,ij}$

Example 1: A parameter space of means of a p-dimensional Normal distribution with constant covariance matrix Σ.

The fundamental tensor of the metric is the inverse of Σ. It is obvious that this space is Euclidean since the fundamental tensor is constant and it may be diagonalized by simple transformation. Of course, K is zero in this case.

Example 2: The trinomial distribution

Let nT_n have a trinomial distribution. The parameter space, D, is given by:

$$D = \{\begin{pmatrix}\theta^1\\ \theta^2\end{pmatrix}: \quad \theta^i>0 \text{ and } \theta^1+\theta^2<1\}.$$

Standard theory implies that

$$\mathcal{L}[\sqrt{n}(T_n-\theta)] \to N(0,\Sigma(\theta))$$

where

$$\Sigma(\theta) = \begin{pmatrix} \theta^1(1-\theta^1) & -\theta^1\theta^2 \\ -\theta^1\theta^2 & \theta^2(1-\theta^2) \end{pmatrix}$$

Therefore, the fundamental tensor of the space is given as follows:

$$(g_{ij}) = \Sigma(\theta)^{-1} = \begin{pmatrix} \dfrac{1-\theta^2}{\theta^1(1-\theta^1-\theta^2)} & \dfrac{1}{1-\theta^1-\theta^2} \\ \dfrac{1}{1-\theta^1-\theta^2} & \dfrac{1-\theta^1}{\theta^2(1-\theta^1-\theta^2)} \end{pmatrix}$$

Holland [1971] showed that there exists no covariance stabilizing transformation in this case and Yoshizawa [1962] obtained the Gaussian curvature as

$$K = \frac{1}{4}\ .$$

This space has a positive constant Gaussian curvature and may be regarded as a sphere in a three dimensional Euclidean space. The coordinates of the Euclidean space are obtained as a set of solutions as follows:

$$(22) \qquad z^1 = 2\sqrt{\theta^1},\ z^2 = 2\sqrt{\theta^2},\ z^3 = 2\sqrt{\theta^3},$$

where $\theta^3 = 1-\theta^1-\theta^2$ (See Appendix B). From (20) the equation of the sphere is

$$(23) \qquad (z^1)^2 + (z^2)^2 + (z^3)^2 = 4,$$

that is,

$$\theta^1 + \theta^2 + \theta^3 = 1.$$

This fact gives a convenient interpretation of the parameter space of a trinomial distribution.

The results obtained here can be extended to the parameter space of the multinomial distribution (See Appendix B).

Example 3: The Normal distribution $N(\mu,\sigma^2)$

Let θ^1 and θ^2 be the mean μ and variance σ^2 respectively. The fundamental tensor of the metric is given as follows:

$$g_{11} = \frac{1}{\theta^2},\ g_{12} = g_{21} = 0,\ g_{22} = \frac{1}{2(\theta^2)^2}$$

The only one independent components of the Riemann-Christoffel curvature R_{1212} is

$$R_{1212} = -\frac{1}{4(\theta^2)^3},$$

and Gaussian curvature is

$$K = \frac{R_{1212}}{g_{11}g_{22}-g_{12}g_{21}} = -\frac{1}{2}.$$

Therefore no covariance stabilizing transformation exists.

Since this space has a negative constant Gaussian curvature it may be regarded as a hyperbolic surface in a three dimensional Euclidean space. The coordinates of the Euclidean space are given as

$$z^1 = \frac{\theta^1}{\sqrt{\theta^2}} = \frac{\mu}{\sigma},$$

(24) $$z^2 = \frac{(\theta^1)^2-2}{2\sqrt{2}\sqrt{\theta^2}} + \frac{\sqrt{\theta^2}}{\sqrt{2}} = \frac{\mu^2-2}{2\sqrt{2}\sigma} + \frac{\sigma}{\sqrt{2}},$$

$$z^3 = \frac{(\theta^1)^2+2}{2\sqrt{2}\sqrt{\theta^2}} + \frac{\sqrt{\theta^2}}{\sqrt{2}} = \frac{\mu^2+2}{2\sqrt{2}\sigma} + \frac{\sigma}{\sqrt{2}},$$

solving the equations (21) and comparing the fundamental forms of the Euclidean space and of the Riemannian space.

The equation of the hyperbolic surface is given by

(25) $$(z^1)^2 + (z^2)^2 - (z^3)^2 = -2.$$

5. Distance and geodesics

The distance between two points in a parameter space has a meaning as a kind of measure for the difference or divergence between two distributions. The distance may be naturally defined as the arc of geodesic curve between two points due to the theory of Riemannian geometry. The elements of length ds is defined by (9) and the arc of geodesic curve is given by the solution of the following equations:

$$\frac{d^2\theta^\ell}{ds^2} + \left\{{\ell \atop jk}\right\} \frac{d\theta^j}{ds}\frac{d\theta^k}{ds} = 0. \tag{26}$$

These equations are Euler's equations of the integral

$$s = \int_{t_1}^{t} \sqrt{g_{ij} \frac{d\theta^i}{dt}\frac{d\theta^j}{dt}}\, dt$$

where t is a parameter which defines real curve. Along the geodesic we have

$$g_{ij} \frac{d\theta^i}{ds}\frac{d\theta^j}{ds} = 1. \tag{27}$$

If we put new coordinates associated with the geodesic passing through the particular point θ_o such that

$$\theta'^i = \left(\frac{d\theta^i}{ds}\right)_o s \tag{28}$$

the arc of the geodesic may be expressed as

$$s^2 = (g_{ij})_o\, \theta'^i \theta'^j . \tag{29}$$

Notice that this form may be interpreted as an extension of Mahalanobis' generalized distance. Moreover if we transform θ'^i to θ''^i such that θ''^1 will express the arc of the geodesic, the fundamental form of θ''^i is reduced to

$$\phi = (d\theta''^1)^2 + g''_{\alpha\beta} d\theta''^\alpha d\theta''^\beta \quad (\alpha,\beta=2,\ldots,p) \tag{30}$$

(See Eisenhart [1926], p. 57). From the form (30) it is seen that if we use the transformation θ''^1 we may get a transformation of random variables which obeys asymptotical Normal distribution with unit variance. Therefore

$$ns^2(x_n:\theta_o)$$

will obey Chi-square distribution of one degree of freedom asymptotically. Then we have

<u>Theorem 2</u>: Under the assumption (1) substituting θ_o in θ

$$ns^2(x_n:\theta_o) \tag{31}$$

asymptotically obeys Chi-square distribution of one degree of freedom, where $s(\theta:\theta_o)$ is the geodesic given as the solution (26) using the inverse of the asymptotic variance of $\sqrt{n}\, x_n$ as a fundamental tensor.

It is known that the equation (27) may reduce to

$$\frac{d^2 z^\alpha}{ds^2} = - Kz^\alpha \tag{32}$$

with respect to the coordinates z^α of the Euclidean space in which a Riemannian space with a constant Gaussian curvature is immersed. There are two cases to be considered according to the sign of K.

1°. K>0. In this case we have

$$\sum_\alpha c_\alpha (z^\alpha - z_o^\alpha)^2 = \frac{4}{K} \sin^2 \frac{\sqrt{K}s}{2}, \tag{33}$$

where z_o^α is the value of z^α at a point θ_o. From (20) we may rewrite (33) as

$$K\Sigma c_\alpha z^\alpha z_o^\alpha = \cos\sqrt{K}\, s. \tag{34}$$

2°. K<0. Similarly we have

$$\sum_\alpha c_\alpha (z^\alpha - z_o^\alpha)^2 = -\frac{4}{K} \sinh^2 \frac{\sqrt{-K}s}{2}, \tag{35}$$

$$K\sum_\alpha c_\alpha z^\alpha z_o^\alpha = \cosh\sqrt{-K}s\ . \tag{36}$$

Example 4: The trinomial distribution of the example 2.

From the example 2, the arc of geodesic is given by:

$$s = 2\cos^{-1}\left(\sqrt{\theta^1\theta_o^1} + \sqrt{\theta^2\theta_o^2} + \sqrt{\theta^3\theta_o^3}\right), \tag{37}$$

substituting (22) and K = 1/4 in 34. In the case of multinomial distribution we have

$$s = 2\cos^{-1}\sum_\alpha \sqrt{\theta^\alpha\theta_o^\alpha} \tag{38}$$

The half of s has been introduced as a measure of divergence by Bhattacharyya [1942].

From theorem 2 it is seen that the asymptotic variance of

(39) $$2\cos^{-1}\sum_{\alpha}\sqrt{\hat{\theta}^{\alpha}\theta^{\alpha}}$$

is 1, where $\hat{\theta}^{\alpha}$ is the usual maximum likelihood estimator of θ^{α}.

Example 5: The Normal distribution of the example 3

Since the Gaussian curvature K is negative constant -1/2, substituting (24) in (33), we obtain the arc of geodesic as follows:

(40) $$s = \sqrt{2}\cosh^{-1}\frac{(\mu-\mu_o)^2+2(\sigma^2+\sigma_o^2)}{4\sigma\sigma_o}.$$

The statistic associated with the geodesic is

(41) $$\sqrt{n}s(\bar{x}_n,S_n^2:\mu_o,\sigma_o^2) = \sqrt{2n}\cosh^{-1}\frac{(\bar{x}-\mu_o)^2+2(S_n^2+\sigma_o^2)}{4\sqrt{S_n^2}\,\sigma_o}$$

and by the theorem 2 it has asymptotic unit variance, where $\bar{x}_n$ and S_n^2 are the sample mean and variance respectively.

6. Fisher's information

Fisher's information matrix in several parameters are defined by:

(42) $$I^{ij} = E\left(\frac{\partial\log L}{\partial\theta^i}\frac{\partial\log L}{\partial\theta^j}\right)$$

where L is the likelihood function of parameters θ^i (Fisher [1921]). It is easily shown that I^{ij} is a covariant tensor under transformations of parameters. Rao [1945] and

Yoshizawa [1962] used I^{ij} as the fundamental tensor of a parameter space. The inverse of the matrix I^{ij} gives the lower bound of the variance of the estimators θ under some regularity condition by Rao-Cramer inequality. The fundamental tensors of the metric in the examples here are quite the same as Fisher's information matrix and they are rather easily calculated directly from (42) or from the following formula equivalent to (42) in the case that maximum likelihood estimators are used:

$$I^{ij} = -E\left(\frac{\partial^2 \log L}{\partial\theta^i \partial\theta^j}\right)$$

Anyway this geometry of parameter spaces is concerned with asymptotic variance and the limit of this method should be considered in this point.

Finally notice that the various definitions of distance or divergence between two distributions, e.g. Kullback's divergence [1959] and Matusita's [1955], are often locally equivalent to Fisher's information.

References

Bhattacharyya, A. (1942) "On a measure of divergence between two multinomial populations" Sankhyā, Vol. 7, 401-6.

Dempster, A.P. (1969) Continuous Multivariate Analysis Addison-Wesley, Reading, Massachusetts.

Eisenhart, L.P. (1926) Riemann Geometry, Princeton University Press, Princeton.

Eisenhart, L.P. (1940) An Introduction to Differential Geometry with Use of the Tensor Calculus, Princeton University Press, Princeton.

Fisher, R.A. (1921) "On the mathematical foundations of theoretical statistics", Phil. Trans. Roy. Soc. A, 222, 309-368.

Holland, P.W. (1971) "On the possibility of a multivariate extension of the variance stabilizing transformations", Memorandum NS-143 Harvard University, Cambridge, Massachusetts.

Kullback, S. (1959) Information Theory and Statistics, John Wiley & Sons, New York.

Mahalanobis, P.C. (1936) "On the generalized distance in Statistics", Proc. Nat. Inst. Sc. Ind., 2, 49-55.

Matusita, K. (1955) "Decision rules, based on the distance, for problems of fit, two samples, and estimation" Ann. Math. Stat., Vol. 26, pp. 631-640.

Rao, C.R. (1945) "Information and the accuracy attainable in the estimation of statistical parameters", Bulletin of the Calcutta Math. Soc., Vol. 37, 81-91.

Veblen, O. (1933) Invariants of Quadratic Differential Forms, Cambridge University Press, London.

Yoshizawa, T. (1962) "Information-distances in parameter spaces" (in Japanese). Unpublished graduation thesis at University of Tokyo.

Appendix A: A proof of the equivalence between Holland's theorem and the theorem 1 in this paper.

Holland's condition for existence of a covariance stabilizing transformation in two dimensional case is

$$(\text{A-1})\quad \frac{\partial}{\partial\theta^2}\left[\frac{1}{\det(A)}\left\{a_{11}\left(\frac{\partial a_{11}}{\partial\theta^2}-\frac{\partial a_{12}}{\partial\theta^1}\right)+a_{21}\left(\frac{\partial a_{21}}{\partial\theta^2}-\frac{\partial a_{22}}{\partial\theta^1}\right)\right\}\right]$$

$$=\frac{\partial}{\partial\theta^1}\left[\frac{1}{\det(A)}\left\{a_{12}\left(\frac{\partial a_{11}}{\partial\theta^2}-\frac{\partial a_{12}}{\partial\theta^1}\right)+a_{22}\left(\frac{\partial a_{21}}{\partial\theta^2}-\frac{\partial a_{22}}{\partial\theta^1}\right)\right\}\right],$$

where A is any matrix (2x2) that satisfies

$$(\text{A-2})\qquad A'A=\Sigma(\theta)^{-1}$$

Our condition (12) is reduced to

$$(\text{A-3})\qquad K=\frac{R_{1212}}{g}=0$$

in two dimensional case and Gaussian curvature K may be written as follows:

$$(\text{A-4})\quad K=\frac{1}{2\sqrt{g}}\left[\frac{\partial}{\partial\theta^1}\left(\frac{g_{12}}{g_{11}\sqrt{g}}\frac{\partial g_{11}}{\partial\theta^2}-\frac{1}{\sqrt{g}}\frac{\partial g_{22}}{\partial\theta^1}\right)\right.$$

$$\left.+\frac{\partial}{\partial\theta^2}\left(\frac{2}{\sqrt{g}}\frac{\partial g_{12}}{\partial\theta^1}-\frac{1}{\sqrt{g}}\frac{\partial g_{11}}{\partial\theta^2}-\frac{g_{12}}{g_{11}\sqrt{g}}\frac{\partial g_{11}}{\partial\theta^1}\right)\right],$$

(See Eisenhart (1940), p. 154 Ex. 10) We will prove that (A-4) is equivalent to (A-3) using the form (A-4).

We may assume that $A(\theta)$ is an upper triangular matrix without loosing generality. Then the components of A are expressed as follows:

$$a_{11} = \sqrt{g_{11}} \,,$$

(A-5)
$$a_{12} = \frac{g_{12}}{\sqrt{g_{11}}} \,, \quad a_{21} = 0,$$

$$a_{22} = \sqrt{g_{22} - \frac{g_{12}^2}{g_{11}}} = \sqrt{\frac{g}{g_{11}}} \,.$$

Partially differentiating the both sides of (A-5), we obtain

$$\frac{\partial a_{11}}{\partial \theta} = \frac{1}{2\sqrt{g_{11}}} \frac{\partial g_{11}}{\partial \theta} \,,$$

(A-6)
$$\frac{\partial a_{12}}{\partial \theta} = \frac{1}{\sqrt{g_{11}}} \frac{\partial g_{12}}{\partial \theta} - \frac{1}{2} \frac{g_{12}}{g_{11}\sqrt{g_{11}}} \frac{\partial g_{11}}{\partial \theta} \,,$$

$$\frac{\partial a_{22}}{\partial \theta} = \frac{1}{2}\sqrt{\frac{g_{11}}{g}} \left(\frac{\partial g_{22}}{\partial \theta} - \frac{2g_{12}}{g_{11}} \frac{\partial g_{12}}{\partial \theta} + \left(\frac{g_{12}}{g_{11}}\right)^2 \frac{\partial g_{11}}{\partial \theta} \right) \,,$$

where θ may take either θ^1 or θ^2. Substituting (A-5) and (A-6) in (A-1), and using $\det(A) = \sqrt{g}$ we rewrite (A-1) as

$$\frac{\partial}{\partial\theta^2}\left[\frac{1}{\sqrt{g}}\sqrt{g_{11}}\left(\frac{1}{2\sqrt{g_{11}}}\frac{\partial g_{11}}{\partial\theta^2}-\frac{1}{\sqrt{g_{11}}}\frac{\partial g_{12}}{\partial\theta^1}+\frac{g_{12}}{2g_{11}\sqrt{g_{11}}}\frac{\partial g_{11}}{\partial\theta^1}\right)\right.$$

$$=\frac{\partial}{\partial\theta^1}\left[\frac{1}{\sqrt{g}}\left\{\frac{g_{12}}{\sqrt{g_{11}}}\left(\frac{1}{2\sqrt{g_{11}}}\frac{\partial g_{11}}{\partial\theta^2}-\frac{1}{\sqrt{g_{11}}}\frac{\partial g_{12}}{\partial\theta^1}+\frac{1}{2}\frac{g_{12}}{g_{11}\sqrt{g_{11}}}\frac{\partial g_{11}}{\partial\theta^1}\right)\right.\right.$$

$$\left.\left.-\sqrt{\frac{g}{g_{11}}}\,\frac{1}{2}\sqrt{\frac{g_{11}}{g}}\left(\frac{\partial g_{22}}{\partial\theta^1}-\frac{2g_{12}}{g_{11}}\frac{\partial g_{12}}{\partial\theta^1}+\left(\frac{g_{12}}{g_{11}}\right)^2\frac{\partial g_{11}}{\partial\theta^1}\right)\right\}\right]$$

After simple algebra we get

$$-\frac{\partial}{\partial\theta^2}\left[\frac{1}{\sqrt{g}}\left(\frac{\partial g_{12}}{\partial\theta^1}-\frac{\partial g_{11}}{2\partial\theta^2}-\frac{g_{12}}{2g_{11}}\frac{\partial g_{11}}{\partial\theta^1}\right)\right]$$

$$=\frac{\partial}{\partial\theta^1}\left[\frac{1}{2\sqrt{g}}\left(\frac{g_{12}}{g_{11}}\frac{\partial g_{11}}{\partial\theta^2}-\frac{\partial g_{22}}{\partial\theta^1}\right)\right]$$

It is easily seen that the above equation is equal to $K=0$.

Appendix B: The multinomial distribution

Let the likelihood of a multinomial distribution be

(B-1)
$$L = \frac{n!}{\prod_{i=1}^{p+1} x_i!} \prod_{i=1}^{p+1} \theta_i^{x_i} ,$$

where θ's are p+1 parameters and x's are random variables. Notice that we use subscripts for parameters in order not to confuse superscripts and squares, etc. The first p parameters are used for the coordinates of the population space since

(B-2)
$$\Sigma\theta_i = 1$$

1°. Fundamental tensor of the metric.

By calculating Fisher's information matrix, the fundamental tensor of the metric is given as below:

(B-3)
$$(g_{ij}) = \begin{pmatrix} \frac{\theta_1+\theta_{p+1}}{\theta_1\theta_{p+1}} & \frac{1}{\theta_{p+1}} & \frac{1}{\theta_{p+1}} & \cdots & \frac{1}{\theta_{p+1}} \\ \frac{1}{\theta_{p+1}} & \frac{\theta_1+\theta_{p+1}}{\theta_2\theta_{p+1}} & \frac{1}{\theta_{p+1}} & \cdots & \frac{1}{\theta_{p+1}} \\ \vdots & & \ddots & & \vdots \\ \frac{1}{\theta_{p+1}} & \frac{1}{\theta_{p+1}} & \cdots & & \frac{\theta_p+\theta_{p+1}}{\theta_p\theta_{p+1}} \end{pmatrix}$$

The covariance matrix of the usual estimators x_i/n, i.e., the inverse of (g_{ij}) is

$$(B\text{-}4)\qquad (g^{ij}) = \begin{pmatrix} \theta_1(1-\theta_1) & -\theta_1\theta_2 & \cdots & -\theta_1\theta_p \\ -\theta_1\theta_2 & \theta_2(1-\theta_2) & \cdots & -\theta_2\theta_p \\ \vdots & & \ddots & \vdots \\ -\theta_p\theta_1 & -\theta_p\theta_2 & \cdots & \theta_p(1-\theta_p) \end{pmatrix}.$$

2°. Christoffel's 3 index symbols.

From (B-3) Partial derivatives of g_{ij} are

$$(B\text{-}5)\qquad \frac{\partial g_{ij}}{\partial \theta_k} = \begin{cases} \dfrac{1}{\theta_{p+1}^2} - \dfrac{1}{\theta_i^2}, & i=j=k, \\[2ex] \dfrac{1}{\theta_{p+1}^2}, & \text{otherwise,} \end{cases}$$

From the definition (14) and (15) Christoffel's 3 index symbols of the first kind are

$$(B\text{-}6)\qquad [ij,k] = \begin{cases} \frac{1}{2}\left(-\dfrac{1}{\theta_i^2} + \dfrac{1}{\theta_{p+1}^2}\right), & i=j=k, \\[2ex] \frac{1}{2}\,\dfrac{1}{\theta_{p+1}^2}, & \text{otherwise,} \end{cases}$$

and Christoffel's 3 index symbols of the second kind are

$$(\text{B-7}) \qquad \left\{ {\ell \atop ij} \right\} = \begin{cases} \dfrac{\theta_\ell}{2\theta_{p+1}}, & i \neq j, \\ \dfrac{\theta_\ell}{2}\left(\dfrac{1}{\theta_i^2} + \dfrac{1}{\theta_{p+1}}\right), & i = j, \quad \neq i, \\ \dfrac{\theta_i}{2\theta_{p+1}} - \dfrac{1-\theta_i}{2\theta_i}, & i = j = \ell\ . \end{cases}$$

3°. Riemann-Christoffel curvature tensor

In two dimensional case, i.e. in the case of the trinomial distribution, Riemann-Christoffel curvature tensor is from (13)

$$(\text{B-8}) \quad R_{1212} = \frac{\partial}{\partial\theta_1}[22,1] - \frac{\partial}{\partial\theta_2}[21,1] + \left\{ {\ell \atop 21} \right\}[12,\ell] - \left\{ {\ell \atop 22} \right\}[11,\ell]$$

$$= \frac{1}{\theta_3^3} - \frac{1}{\theta_3^3} + \frac{\theta_1}{2\theta_3}\frac{L}{2\theta_3^2} + \frac{\theta_2}{2\theta_3}\frac{1}{2\theta}$$

$$- \frac{\theta_1}{2}\left(\frac{1}{\theta_2} + \frac{1}{\theta_3}\right)\frac{1}{2}\left(-\frac{1}{\theta_1^2} + \frac{1}{\theta_3^2}\right) - \left(\frac{\theta_2}{2\theta_3} - \frac{1-\theta_2}{2\theta_2}\right)\frac{1}{2\theta_3^2}$$

$$= \frac{1}{4\theta_1\theta_2\theta_2}\ .$$

Therefore

$$(\text{B-9}) \qquad K = \frac{R_{1212}}{g} = \frac{1}{4}\ .$$

In p dimensional case it is seen from the identities (16) that R_{hijk} is equal to zero if $h = i$ or $j = k$. Under the condition $h \neq i$ and $j \neq k$, noticing that

$$\frac{\partial}{\partial\theta_j}[ik,h] - \frac{\partial}{\partial\theta_k}[ij,h] = 0$$

R_{hijk}'s are calculated as follows:

$$(B\text{-}10)\quad \begin{cases} \circ\ h=j,\ i\neq k \\ \quad R_{hihk} = \dfrac{1}{4\theta_{p+1}\theta_h} \\ \circ\ h\neq j,\ i=k \\ \quad R_{hiji} = R_{ihij} = \dfrac{1}{4\theta_{p+1}\theta_i} \\ \circ\ h=j,\ i=k \\ \quad R_{hihi} = \dfrac{1}{4}\{(\dfrac{1}{\theta_i} + \dfrac{1}{\theta_{p+1}})(\dfrac{1}{\theta_h} + \dfrac{1}{\theta_{p+1}}) - \dfrac{1}{\theta_{p+1}^2}\} \end{cases}$$

Therefore it is easily seen that Gaussian curvature

$$(B\text{-}11)\qquad K = \frac{R_{hijk}}{g_{hj}g_{ik}-g_{hk}g_{ij}} = \frac{1}{4}\ .$$

$\underline{4}$°. Coordinates of a Euclidean space in which the population space of multinomial distribution is immersed.

Let z^α denote the coordinates of the Euclidean space. z^α's are given as a set of solutions of (21), i.e.

$$\frac{\partial^2 z^\alpha}{\partial\theta_i\partial\theta_j} - \sum_{h=1}^{p}\frac{\partial z^\alpha}{\partial\theta_h}\left\{{h \atop ij}\right\} = -\frac{1}{4}\, g_{ij}z^\alpha\ .$$

It is easily seen that a set of solutions is given as

$$z^\alpha = c\sqrt{\theta_\alpha}\ ,\quad \alpha=1,\dots,p+1.$$

Appendix C: The normal distribution $N(\mu, \sigma^2)$

Let θ^1 and θ^2 be μ and σ^2 respectively. We shall use θ_1 and θ_2 for θ^1 and θ^2 because of practical convenience in algebra.

1°. Fundamental tensor of the metric

Standard theory implies that

(C-1) $$\Sigma(\theta) = \begin{pmatrix} \theta_2 & 0 \\ 0 & 2\theta_2^2 \end{pmatrix} = (g^{ij}) .$$

Therefore the fundamental tensor of the metric is

(C-2) $$g_{11} = \frac{1}{\theta_2}, \quad g_{12} = g_{21} = 0, \quad g_{22} = \frac{1}{2\theta_2^2} .$$

2°. Christoffel's 3 index symbols.

Since partial derivatives by θ_1 and θ_2 are zeros except that

(C-3) $$\frac{\partial g_{11}}{\partial \theta_2} = -\frac{1}{\theta_2^2}, \text{ and } \frac{\partial g_{22}}{\partial \theta_2} = -\frac{1}{\theta_2^3} ,$$

from (14) and (15) Christoffel's 3 index symbols of the first and second kinds are as follows:

First kind:

(C-4)
$$[11,1] = 0,$$
$$[11.2] = \frac{1}{2\theta_2^2},$$
$$[12,1] = [21,1] = -\frac{1}{2\theta_2^2},$$
$$[12,2] = [21,2] = 0,$$
$$[21,1] = 0,$$
$$[22,2] = \frac{-1}{2\theta_2^3},$$

Second kind:

$$
\begin{aligned}
&\left\{{1 \atop 11}\right\} = 0,\\
&\left\{{2 \atop 11}\right\} = 1,\\
\text{(C-5)}\qquad &\left\{{1 \atop 21}\right\} = \left\{{1 \atop 21}\right\} = -\frac{1}{2\theta_2},\\
&\left\{{2 \atop 12}\right\} = \left\{{2 \atop 21}\right\} = 0,\\
&\left\{{1 \atop 22}\right\} = 0,\\
&\left\{{2 \atop 22}\right\} = -\frac{1}{\theta_2}.
\end{aligned}
$$

3°. Riemann-Christoffel curvature tensor

Substituting (C-4) and (C-5) in (13), we obtain

$$
\begin{aligned}
R_{1212} &= \frac{\partial}{\partial\theta_1}[22,1] - \frac{\partial}{\partial\theta_2}[21,1]\\
&\quad + \left\{{1 \atop 21}\right\}[12,1] + \left\{{2 \atop 21}\right\}[12,2] - \left\{{1 \atop 22}\right\}[11,1] - \left\{{2 \atop 22}\right\}[11,2]\\
&= -\frac{1}{\theta_2^3} + \left(\frac{-1}{2\theta_2}\right)\left(\frac{-1}{2\theta_2^2}\right) - \left(\frac{1}{\theta_2}\right)\left(\frac{1}{2\theta_2^3}\right)\\
&= -\frac{1}{4\theta_2^3}.
\end{aligned}
$$

Therefore no covariance stabilizing transformation exists. Gaussian curvature becomes

$$K = \frac{R_{1212}}{g} = -\frac{1}{2},$$

where g is the determinant of (g_{ij}).

4°. Coordinates of a Euclidean space in which the population space is immersed.

Let z^α denote the coordinates of the Euclidean space. From (21) z^α must satisfy the following equations:

$$\begin{cases} \dfrac{\partial^2 z^\alpha}{\partial \theta_1^2} - \dfrac{\partial z^\alpha}{\partial \theta_2} = \dfrac{z^\alpha}{2\theta_2}, \\[2ex] \dfrac{\partial^2 z^\alpha}{\partial \theta_1 \partial \theta_2} + \dfrac{1}{2\theta_2}\dfrac{\partial z^\alpha}{\partial \theta_1} = 0, \\[2ex] \dfrac{\partial^2 z^\alpha}{\partial \theta_2^2} + \dfrac{1}{\theta_2}\dfrac{\partial z^\alpha}{\partial \theta_2} = \dfrac{z^\alpha}{4\theta_2^2} . \end{cases}$$

From the third equation it is seen that z^α must be of the form

$$\frac{f(\theta_1)}{\sqrt{\theta_2}} + g(\theta_1)\sqrt{\theta_2}$$

where f and g are some functions of only θ_1. Substituting it in the first and second equations, it is seen that $f(\theta_1)$ and $g(\theta_1)$ must satisfy

$$\frac{d^2 f}{d\theta_1^2} = g, \quad \frac{dg}{d\theta_1} = 0.$$

Therefore z^α must be of the form

$$\frac{a_\alpha \theta_1^2 + b_\alpha \theta_1 + d_\alpha}{\sqrt{\theta_2}} + 2a_\alpha \sqrt{\theta_2}.$$

Since the fundamental form of the Euclidean space is of the form

$$c_1(dz^1)^2 + c_2(dz^2)^2 + c_3(dz^3)^2$$

where c_1, c_2 and c_3 are plus or minus one, comparing it with the fundamental form of the Riemannian space

$$\frac{1}{\theta_2}(d\theta_1)^2 + \frac{1}{2\theta_2^2}(d\theta_2)^2$$

we obtain that $c_1=c_2=-c_3=1$ and that

$$z^1 = \frac{\theta_1}{\sqrt{\theta_2}} ,$$

$$z^2 = \frac{\theta_1^2-2}{2\sqrt{2}\sqrt{\theta_2}} + \frac{\sqrt{\theta_2}}{\sqrt{2}} ,$$

$$z^3 = \frac{\theta_1^2+2}{2\sqrt{2}\sqrt{\theta_2}} + \frac{\sqrt{\theta_2}}{\sqrt{2}} .$$

It is easily seen that the coordinates satisfy (20), i.e.

$$(z^1)^2+(z^2)^2-(z^3)^2 = \frac{1}{K} = -2 .$$

5°. Transformation by geodesic

It is seen from (24) that the arc of geodesic from (μ_o,σ_o^2) to (μ,σ^2) is

$$s = \sqrt{2}\cosh^{-1}\frac{(\mu-\mu_o)+2(\sigma^2+\sigma_o^2)}{4\sigma\sigma_o} .$$

By the law of transformation (5), if we use s as a transformation it follows that

$$g'^{11} = \sigma^2 \left(\frac{\partial s}{\partial \mu}\right)^2 + 2\sigma^4 \left(\frac{\partial s}{\partial(\sigma^2)}\right)^2$$

$$= \sigma^2 \left(\frac{\partial s}{\partial \mu}\right)^2 + \frac{\sigma^2}{2} \left(\frac{\partial s}{\partial \sigma}\right)^2 .$$

Since partial derivatives s by μ and σ are

$$\frac{\partial s}{\partial \mu} = \sqrt{2}\,\frac{1}{\sinh\frac{s}{\sqrt{2}}} \cdot \frac{2(\mu-\mu_o)}{4\sigma\sigma_o} ,$$

$$\frac{\partial s}{\partial \sigma} = \sqrt{2}\,\frac{1}{\sinh\frac{s}{\sqrt{2}}}\,\frac{4\sigma^2-[(\mu-\mu_o)^2+2(\sigma^2+\sigma_o^2)]}{4\sigma_o\sigma^2} ,$$

it is seen that

$$g'^{11} = 1.$$

This fact shows that theorem 2 is certainly valid in this case.